iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** III **Month of publication:** March 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

A Novel Semiautomatic Crowdsourcing Predictors for Faster Statistical Analysis Based Upon User Inputs

Gaurav S. Barde[1], Ramesh S. Dhavane[1], Pavankumar K. Khole[1], Varun P. Mehta[1], Prof. Arindam Dasgupta[2]

[1,2]Department of Information Technology, Amrutvahini College of Engineering, Sangamner, Maharashtra, India.

Abstract: ***Creating models from multiple data sets and deciding which data set is to be mined is now a day become more and more automated. But for selecting required data we required knowledge and experience, usually provided by domain experts. This paper gives the new approach to machine science by which first time non domain expert can create methodologies and provide values to those methodologies so that they can be useful for predicting behavioral consequence of interest. This is achieved by building a web platform in which a group of people interact with each other to give answers to questions as well as to predict behavioral consequences and to present a question to their peers. This gives a continuously improvising online survey and leads to predict the behavioral consequences of user with the help of their responses to survey questions formed by the user. Here we explain two web-based approaches to this concept: the first website predicts the user's daily electricity consumption and other predicts body mass index. As daily increase in use of web this website gives large outputs in future.***

Keywords: Crowdsourcing, machine science, surveys, social media, human behavior modeling

I. INTRODUCTION

Now days there are many problems in which exact solution is not possible. In such cases one has to predict the consecutive result. In such problems a team of experts is required for each individual domain which results in excess loss of human efforts. For example, the survey designer must be an expert of that domain to choose appropriate questions related to respected domain. An engineer must keep correlation and well known approach of design in order to judge which concept will be more efficient such a way that it will increase the performance.

Necessity of domain expert is the main drawback of this approach. However, using the knowledge of crowd to understand the difficult problems will harness the effectiveness of result. Thus, the goal of this web-based approach is to achieve active participation of crowd in suggesting questions along with providing answers to the given questions which leads to development of predictive model.

A. Machine Science

Machine science consists of automation of as many scientific concepts as possible. But in case of machine science it is very difficult to decide which variable of subset is to be selected. It is also very difficult to decide which variable to be automated. When we discuss about the prediction problem machine science, sometimes is unable to select the variable which can predict the outcome of interest.

In this paper we describe a method by which a non-domain expert can able to generate the variables for successful

ISSN: 2321-9653

modeling. Let's we see it in more, user come to this website in which behavioral outcomes (like daily electricity consumption and body mass index) is modeled. The user can give their own outcomes (like their own body mass index). Then user can give the answers for questions which can be a predictive of those consequences. The model is get generated as the data set is get increased. User may suggest a question which when answered by others becomes the new independent variable.

B. Crowdsourcing

The fast growing user generated data on Internet is an example of crowdsourcing which is very helpful where previously a team of experts is needed. Harnessing the experience and effort of large numbers of individuals is known as "crowdsourcing" and it is now a day play a vital role in many fields and researches.

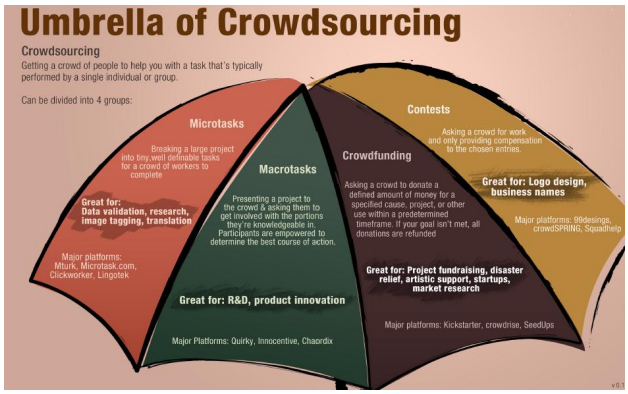


Fig. 1 Umbrella of Crowdsourcing

Above diagram gives us the idea about how crowdsourcing works and how it is effective in many fields now days. In many cases when we have some problems (like extra BMI or excess electricity consumption), each time we have to go to domain experts for solution of our problem. This problem can be solve with the help of crowdsourcing i.e. user can solve their own problems by their own selves. The best example which proves the effectiveness of crowdsourcing is Amazon's Mechanical Turk. In this one can explain a "Human Intelligence Task" such as characterizing data, transcribing spoken language, or creating data visualizations with the help of group of people which is very difficult for a computer alone.

II. METHODOLOGY

In this paper we describe the application of crowdsourcing in cyber infrastructure such that:

(1) Investigator provides some behavioral outcomes that are modeled.

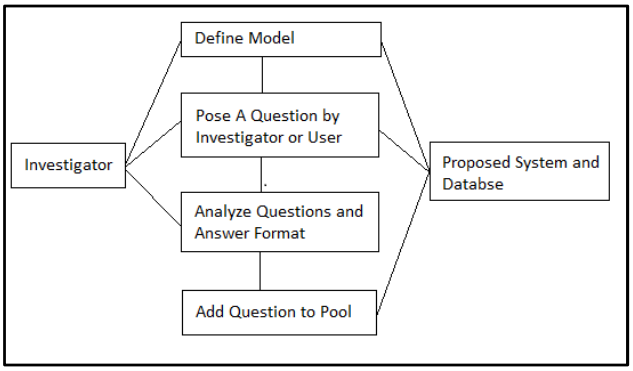


Fig. 2 Investigator Model

In this phase of system flow investigator defines a model for which he wants to predict the consequence then as per requirement he can accept or pose a question. Whatever the accepted question from user the investigator can add it to question pool after analyzing the question if it is suitable for the module.

(2)Data required for our model is collected from user i.e. from human volunteers. They may or may not be domain expert.

If the user is new then he first register himself as a authenticated user then he have to select for which model he wants to predict his behavioral consequence then he can also pose a question which he thinks suitable for the respected module. At last the investigator can generate the user's behavioral consequence depending on the answers given by the user.

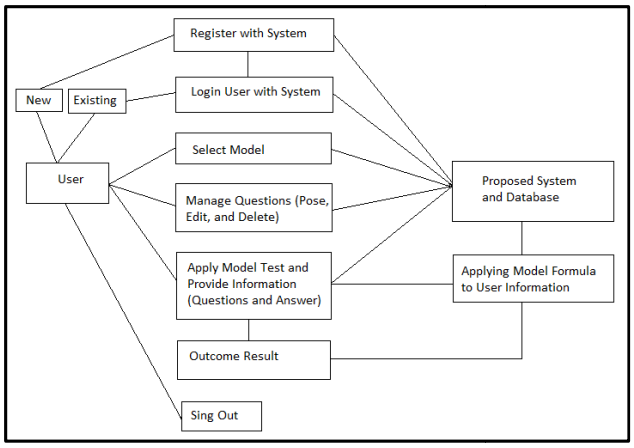


Fig. 3 User Model

III. Benefits And Callenges

In both the cases like body mass index and daily electricity consumption the participants plays a vital role to highlight at least one behavioral consequence. The first major challenge is that what happen if the number of users will provide with the number of questions at a time to the website. The system will get overflow. This problem can be overcome with the help of dynamic filtering of questions. If we can able to filter the questions dynamically it will avoid the overflow of system.

A. User Fatigue

The other challenge of the system is the user fatigue. It may happen that the user answers only a small instance of all questions and due to this some question may get more response than others. As we know that the questions get added to question pool as per the user suggests it. So questions that are present at fist will get the more response than others.

The user may answer the questions that are less predictive than those which are more predictive and it leads to wrong prediction.

B. User Motivation

This is also a challenge in this case as we have to each time motivate the user to answer all the questions in the question pool to generate proper outcome result. If the user will not answer the question whatever the consequences happen that will not provide the proper outcome result. The prediction is totally depends on users response so each time we have to motivate the user.

C. Rare Outcomes

Sometimes it may happen that the user suffering from the rare disease visit the website and if he gives the answers to questions it sometime may lead to wrong prediction of outcomes. All the problems we have discussed above can be overcome by motivating users to give proper answers and also to attempt all the questions.

D. Faster Result

As we have discussed earlier the system is semiautomatic it means we can generate the result in less time and less efforts. Whatever the time that required visiting or communicating with domain expert get saved as the system can generate the result without the help of domain expert.

Conclusion

In closing, this paper has presented a novel contribution to the growing field of machine science in which the formulation of observables for a modeling task and the populating of those observables with values can be offloaded to the human group being modeled.

Aknowledgement

The authors acknowledge with valuable contributions from different reviews, different searches and useful discussions.

References

[1] J. Bongard and H. Lipson, “Automated reverse engineering of nonlinear dynamical systems,” Proceedings of the National Academy of Sciences, vol. 104, no. 24, pp. 9943–9948, 2007.

[2] J. Evans and A. Rzhetsky, "Machine science," Science, vol. 329, no. 5990, p. 399, 2010.

[3] R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," Nature, vol. 427, pp. 247–252, 2004.

[4] R. King, J. Rowland, S. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. Soldatova et al., "The automation of science," Science, vol. 324, no. 5923, p. 85, 2009.

[5] J. Bongard, V. Zykov, and H. Lipson, "Resilient machines through continuous self-modeling," Science, vol. 314, pp. 1118–1121, 2006.

[6] J. Giles, "Internet encyclopedias go head to head," Nature, vol. 438, no. 15, pp. 900–901, 2005.

[7] D. C. Brabham, "Crowdsourcing as a model for problem solving," Convergence, vol. 14, pp. 75–90, 2008.

[8] A. Sorokin and D. Forsyth, "Utility data annotation with amazon mechanical turk," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008.

[9] M. Marge, S. Banerjee, and A. Rudnicky, "Using the amazon mechanical turk for transcription of spoken language," in Proc. IEEE International Conference on Acoustics Speech and Signal Processing, 2010.

[10] N. Kong, J. Heer, and M. Agrawala, "Perceptual guidelines for creating rectangular treemaps," IEEE Transactions on Visualization and Com- puter Graphics, vol. 16, no. 6, 2010.

[11] A. Kittur, E. Chi, and B. Suh, "Crowdsourcing user studies with mechanical turk," in Proc.Twenty-sixth annual SIGCHI conference on human factors in computing systems, 2008.

[12] D. Wightman, "Crowdsourcing human-based computation," in Proceed- ings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, 2010.

[13] B. Fitzgerald, "The transformation of open source software," Manage- ment Information Systems Quarterly, vol. 30, no. 3, pp. 587–598, 2006.

[14] J. Howe, Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business. Crown Business, 2009.

[15] N. Thurman, "Forums for citizen journalists? adoption of user generated content initiatives by online news media," New Media and Society, vol. 10, no. 1, 2008.

[16] C. DiBona, M. Stone, and D. Cooper, Open Source 2.0: The Continuing Evolution. O'Reilly Media, 2005.

[17] J. Leskovec, L. Adamic, and B. Huberman, The Dynamics of Viral Marketing. New York: ACM Press, 2007.

[18] K. Lerman, "Social networks and social information filtering on digg," arXiv: cs/0612046v1, 2006.

[19] D. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, "Seti@home: an experiment in public-resource computing," Communi- cations of the ACM, vol. 45, no. 11, pp. 56–61, 2002.

[20] J. Cohn, "Citizen science: Can volunteers do real research?" BioScience, vol. 58, no. 3, pp. 192–197, 2008.

[21] J. Silvertown, "A new dawn for citizen science," Trends in Ecology & Evolution, vol. 24, no. 9, pp. 467–471, 2009.

[22] A. Beberg, D. Ensign, G. Jayachandran, S. Khaliq, and V. Pande, "Folding@home: Lessons from eight years of volunteer distributed computing," in IEEE International Symposium on Parallel Distributed Processing, May 2009, pp. 1–8.

[23] C. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. Raddick, R. Nichol, A. Szalay, D. Andreescu et al., "Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey?" Monthly Notices of the Royal Astronomical Society, vol. 389, no. 3, pp. 1179–1189, 2008.

[24] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovi´ c et al., "Predicting protein structures with a multiplayer online game," Nature, vol. 466, no. 7307, pp. 756– 760, 2010.

[25] A. Kittur, "Crowdsourcing, collaboration and creativity," XRDS, vol. 17, no. 2, pp. 22–26, 2010.

[26] S. Bowman, S. Gortmaker, C. Ebbeling, M. Pereira, and D. Ludwig, "Effects of fast-food consumption on energy intake and diet quality among children in a national household survey," Pediatrics, vol. 113, no. 1, p. 112, 2004.

[27] J. Currie, S. DellaVigna, E. Moretti, and V. Pathania, "The effect of fast food restaurants on obesity and weight gain," American Economic Journal: Economic Policy, vol. 2, no. 3, pp. 32–63, 2010.

[28] S. Z. Attari, M. L. DeKay, C. I. Davidson, and W. B. de Bruinc, "Public perceptions of energy consumption and savings," Proceedings of the National Academy of Sciences, Aug. 16 2010.

[29] Microsoft. (2011) Microsoft hohm. [Online]. Available: www.microsoft-hohm.com

[30] E. Mills, "The home energy saver: Documentation of calculation methodology, input data, and infrastructure," Lawrence Berkeley National Laboratory, Tech. Rep. LBNL-51938, 2008.

[31] H. Allcott, "Social norms and energy conservation," Journal of Public Economics, 2011.

[32] J. E. Petersen, V. Shunturov, K. Janda, G. Platt, and K. Weinberger, "Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives," International Journal of Sustainability in Higher Education, vol. 8, no. 1, pp. 16–33, 2007.

[33] L. Kaufman, "Utilities turn their customers green, with envy," The New York Times, Jan. 30 2009.

[34] P. Slovic, "Trust, emotion, sex, politics, and science: Surveying the risk- assessment battlefield," Risk Analysis, vol. 19, no. 4, 1999.

[35] G. S. Guthridge, "Understanding consumer preferences in energy efficiency: Accenture end-consumer observatory on electricity management," Accenture, Tech. Rep. ACC10-0229, 2010.

[36] A. Romero-Corral, V. Somers, J. Sierra-Johnson, R. Thomas, M. Collazo-Clavell, J. Korinek, T. Allison, J. Batsis, F. Sert-Kuniyoshi, and F. Lopez-Jimenez, "Accuracy of body mass index in diagnosing obe- sity in the adult general population," International Journal of Obesity, vol. 32, no. 6, pp. 959–966, 2008.

[37] L. Barness, J. Opitz, and E. Gilbert-Barness, "Obesity: genetic, molecu- lar, and environmental aspects," American Journal of Medical Genetics Part A, vol. 143, no. 24, pp. 3016–3034, 2007.

[38] T. Parsons, C. Power, S. Logan, and C. Summerbell, "Childhood predictors of adult obesity: a systematic review." International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity, vol. 23, p. S1, 1999.

[39] Y. Wang and M. Beydoun, "The obesity epidemic in the United States— gender, age, socioeconomic, racial/ethnic, and geographic characteris- tics: a systematic review and meta-regression analysis," Epidemiologic reviews, vol. 29, no. 1, p. 6, 2007.

[40] A. Clauset, C. Rohilla Shalizi, and M. Newman, "Power-law distributions in empirical data," SIAM review, vol. 51, no. 4, pp. 661–703, 2009.

[41] P. Boumtje, C. Huang, J. Lee, and B. Lin, "Dietary habits, demographics, and the development of overweight and obesity among children in the United States," Food Policy, vol. 30, no. 2, pp. 115–128, 2005.

[42] A. Herbert, N. Gerry, M. McQueen, I. Heid, A. Pfeufer, T. Illig, H. Wichmann, T. Meitinger, D. Hunter, F. Hu et al., "A common genetic variant is associated with adult and childhood obesity," Science, vol. 312, no. 5771, p. 279, 2006.

[43] M. Friedman and K. Brownell, "Psychological correlates of obesity: Moving to the next research generation." Psychological Bulletin, vol. 117, no. 1, p. 3, 1995.

[44] M. Van der Merwe, "Psychological correlates of obesity in women," International Journal of Obesity, vol. 31, pp. S14–S18, 2007.

[45] R. Bonow and R. Eckel, "Diet, obesity, and cardiovascular risk," N Engl J Med, vol. 348, no. 21, pp. 2057–2058, 2003.

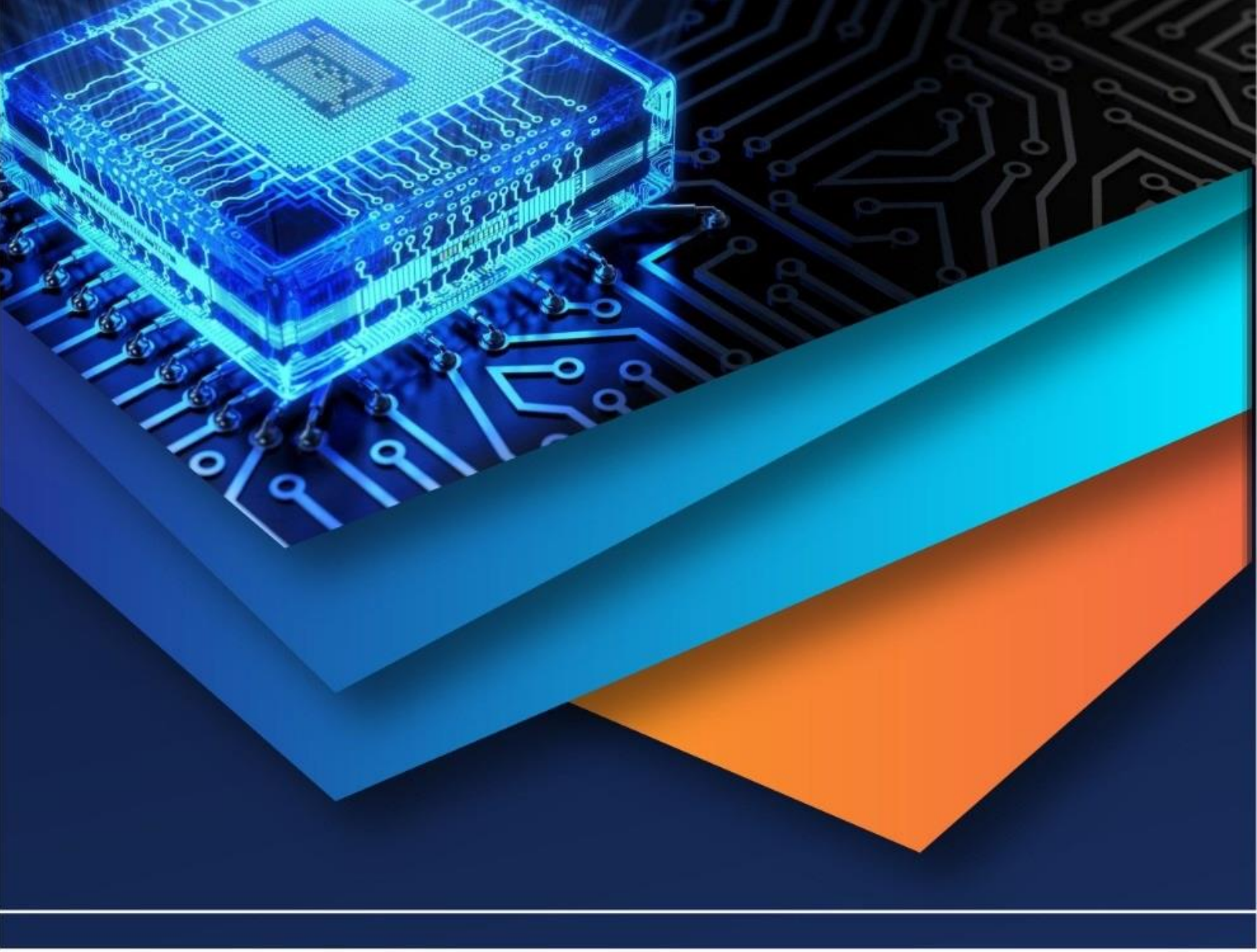
[46] R. Ewing, T. Schmid, R. Killingsworth, A. Zlot, and S. Raudenbush, "Relationship between urban sprawl and physical activity, obesity, and morbidity," Urban Ecology, pp. 567–582, 2008.

[47] S. Grogan, Body image: Understanding body dissatisfaction in men, women, and children. Taylor & Francis, 2008.

[48] M. Skurichina and R. Duin, "Bagging, boosting and the random subspace method for linear classifiers," Pattern Analysis & Applications, vol. 5, no. 2, pp. 121–135, 2002.

[49] Z. Lu, X. Wu, and J. Bongard, "Active learning with adaptive heterogeneous ensembles," in Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on. IEEE, 2009, pp. 327–336.

[50] H. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in Proceedings of the fifth annual workshop on Computational learning theory. ACM, 1992, pp. 287–294.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)